International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 05 | May -2020 ISSN: 2582-3930

PRIORI DATA BASED COVID-19 POOL TESTING TO OVERCOME KIT SHORTAGE

Dr. Indradeep Verma[1], Mr. Akash Kumar Shukla[2], Ms. Aastha Sachan[3], Ms. Namrata Agarwal[4]

[1]*Associate Professor, Department of Computer Science and Engineering, IIMT College of Engineering, Greater Noida, Dr. A.P.J. Abdul Kalam Technical University, Uttar Pradesh*

[2]*U.G. Student, Department of Computer Science and Engineering, IIMT College of Engineering, Greater Noida, Dr. A.P.J. Abdul Kalam Technical University, Uttar Pradesh*

[3]*U.G. Student, Department of Computer Science and Engineering, IIMT College of Engineering, Greater Noida, Dr. A.P.J. Abdul Kalam Technical University, Uttar Pradesh*

[4]*U.G. Student, Department of Computer Science and Engineering, IIMT College of Engineering, Greater Noida, Dr. A.P.J. Abdul Kalam Technical University, Uttar Pradesh*

ABSTRACT

This paper aims at analyzing how technologies, like AI, can enable us to take better measures against COVID-19 which has caused multi-disciplinary changes at the global level.
Governments are seen trying to curb the spread of COVID-19. But one of the major problems they have been facing is the shortage of testing equipment.
Considering this a strategy for finding alternative solutions is a must to ensure minimizing the number of tests that needed to be done. One such approach is: pool sampling, i.e. combined patient samples and testing the combine samples once. Pooling can succeed at a unitary cost, if all the samples taken are negative. But if a single sample comes out to be positive then infected patient does not mean failure.
This paper describes how to optimally detect infected patients in pool samples, i.e. using a minimum number of tests to exactly recognize them, by making an assumption the a priori probabilities that every patient is healthy.

Estimation of those probabilities using questionnaires, supervised machine learning or clinical examinations can be done. The algorithmic results achieved, are like informed divide-and-conquer methodologies and are efficient at performance.

Keywords: COVID-19, AI, Pool Testing, Social-Distancing, Sampling, Corona-Virus

I. Introduction

1. COVID-19 Infection Tests

The **COVID-19** pandemic is seen to be spreading and has significantly impacted the healthcare systems throughout the globe. Although Stay-at-home and social distancing orders are enforced in many countries to curb disease's spread, at the same time also bringing in the major hits to the economic equations and in social structures.
One major reason why social distancing is preferred is the lack of testing tools at the vast level. Rapid detection of people who are infected with COVID-19 and those who have been in their contact is an essential component in controlling the spread of the pandemic. In the case of developed countries like the US, the current estimation is that at least 500,000 Covid-19 tests will need to be performed daily to successfully reopen the economy and make things back running back to normal. Unfortunately, as we are attempting to limit the global COVID-19 infection, even our best efforts are slowed down by the severe shortages of COVID-19 testing kits.
The testing procedure for COVID-19 is usually performed using any one of the following methods:

Molecular Diagnostic Test: This detects the presence of SARS-COV-2 nucleic acids in the blood particles. The presence of SARS-COV2 reflects the presence of the virus in the body and exposure to the infection.

Serological diagnostic tests: This test identifies the presence of antibodies (e.g., IgM, IgG) to SARS-COV-2 in samples taken. The **Serological test**, also known as **Antibody Test**, helps identify not only those who are ill, but also those who have been infected and might have recovered, as antibodies are still present in their blood. This recognition plays an important aspect for numerous purposes.

First, this test can be used to differentiate between sick and people prone to risk. Additionally, capable of identifying people containing antibodies & can be used to carry out research on COVID-19 with their plasma being used as a test sample.

But the facilities of carrying out both tests are in very short supply. Governments have taken several measures to work around this shortage, one such example includes from USA where the FDA4 has become more liberal on approving of COVID-19 tests via the Emergency Use Authorization (EUA), along with an attempts to boost the number of locally produced test kits to reach a throughput of 100,000 kits per day. Those efforts cannot, however, be followed by many countries due to limited industrial resources and technologies and there remains a whole bunch of regions like the Africa continent, Asia and Latin America that are under constant threat.

2. Pool Testing

To optimize the use of available tests which are already in short supply, reduce costs and save time, pool testing can be applied in which multiple samples are mixed, and the resulting '**batch**' is tested using the same amount of resources that would have been required to test one individual sample. The basic objective is to increase the capacity of laboratories to enable them of carrying out more testes for surveillance rather than diagnostic purposes. However, when the presence of any single positive test case fails the test indicating the presence of positive case but no reference of which one. The most basic approach is to test individuals, leading to increased overheads. This context received enormous attention including set up of advisory for addressing the issues.

Such an advisory followed a feasibility study at the **Virus Research & Diagnostic Laboratory** at **King George's Medical University**, **Lucknow**. The study exhibits "executing real-time PCR for COVID-19 by pooling 5 samples of TS/NS (200 microliters/sample) is attainable when the general rates of contamination are less." The ICMR has put an upper limit of **five samples which can be pooled**, this is to steer clear of false negatives because of excessive dilution. More samples can be pooled if the purpose is research only.
It is quite important to distinguish between two types of pool tests that can be performed:

Adaptive tests are those samples of testing depending on prior tested samples and **Nonadaptive** tests, are those where testing is planned. Pool tests are also described as either probabilistic or combinatorial. Concluding, probabilistic models start with a probability distribution over the given sample space and try to optimize the average number of tests required to test all the subjects. On the contrary, combinatorial algorithms focus on minimizing the worst-case figure of tests when the probability distribution governing the experiment is unknown. This paper is in reference to adaptive probabilistic tests.

3. Reference Research Data

Pool testing has been previously used to test large partitions of the mass population (even as a most feasible method, when test availability was quite low) as in the cases to identify viral diseases, such as HIV [NABB19], ZIKA [BMBM17], and INFLUENZA [VMW+12] in the past and outcomes were quite effective to overcome the short supply of kits. Besides, Pool testing has also been suggested as an effective screening method for routine HCV, HBV, and HIV-1 PCR donors for a blood-bank.
In light of the recent pandemic and considering the need of the time, the idea of Pool testing is becoming more and more appealing to implement and effective both at the same time. It is currently the official testing procedure in Israel, Germany, South Korea, and some US and Indian states such as Delhi, Gujarat, Lucknow.

As per WHO, the resolution of testing on clinical and methodological factors that link to an evaluation of the probability of infection. PCR tests of asymptomatic or mildly symptomatic patients may be considered in the evaluation of individuals who may have exposure with a COVID-19 case. Screening protocols should be modified to the current situation. The cases are being regularly inquired and updated as new statistics become available. For the WHO suspect case definition see: Global Supervision for human contamination with coronavirus disease (COVID-2019) (10).

However, Field research focusing on reducing the number of tests did not analyze prior information strategies that could be utilized to collect and analyze for optimization of the test and resources but instead provided simulation (or small sample) results showing the benefits of pool testing which nowhere covers the full potential pool testing can be utilized to. In most of the cases pool tests have been used only to showcase the domains of problem domains that can be solved to achieve optimized screening but not as a standardized approach to differentiate infected v/s healthy.

The work done on pool testing so far in the area of research could be utilized to make pool testing more effective. Few such application source domains are:

1. Assays [Woo20] Project, which consists of pool testing templates comprise of standard 96 well plates. The Origami XL3 design tests comprising of 1120 patients in 94 assay wells, which is a good enough sample set.

2. Yelin et al. [YAST+20] demonstrated that pool testing can be used effectively to identify one positive SARS-COV-2 result within 32 samples, and possibly within 64 samples if the cycles are amplified, with an estimated false-negative rate of 10%. [Täu20] uses a strategy consisting of running 'cross batches', where the same individuals are tested several times but in different pools, which eventually leads to positive sample identification. The resulting approach ends up using more tests overall (since it tests every individual more than once) than the strategy proposed in this work and does not exploit prior information. Similarly, Sinnott-Armstrong et al. [SASKH20] suggested to identify low-risk individuals (i.e. asymptomatic and mild cases) and to test them as a pool using a matrix-based method, to reduce the number of tests required by up to eight-fold, depending on the prevalence. It is assumed that a successful emergency application of the refined pool testing procedures classified in this paper would improve the COVID-19 testing capacity significantly.

4. Contribution

This paper deviates from the above-mentioned approaches and so far used procedures by considering the fact of the availability of extra information i.e. the a priori probability that each given test is negative. In practice, we may either assume that such probabilities are given, estimated from patient trust metrics, or are learned from past COVID-19 tests. We assume in this work that these probabilities are known. We show that it is possible to find positive samples optimally, i.e., by performing on average the minimum number of tests. This approach turns out to be faster than heuristic divide-and-conquer testing that is applied in most of the cases. An outcome of this analysis is the methodological description of a procedure that can be used to design a testing technique that is comparatively faster and cost-effective.

II. Intuition over the approach

Before introducing the mathematical model and its representation it's important to describe the guiding principle of the algorithm. We will present the hypothesis by considering a very small case of three samples. These samples can be tested individually or together like a batch in a pool.

Representing in terms of complexity of an algorithm, individual COVID-19 testing claims a minimum of two units of work—check one sample, then check the other. The approach of Pool-checking requires at least one COVID-19 test. Considering this, there is a high probability that both subjects will turn out to be negative, then pool testing is interesting: If both samples are indeed negative, this way we have halved the COVID-19 test's cost. However, in case of failure, we are back to square one, with a slight overhead because of these samples (at least) is positive, and no certainty of which one and this is an overhead that we are trying to mitigate using prior information.

In this paper, we identify when to check samples individually, and when to pool-check them instead—including all possible generalizations when there are more than 2 samples. We assume that the probability of a sample being positive is known to us a priori. The result is a testing 'meta procedure' that offers the best alternative to sequential and individual testing.

III. Mathematical Modelling of the problem

Testing procedures -
We consider a collection of n samples.
Let [n] denote {1...n}, and $\Omega = P([n]) \setminus \{/0\}$, where P is the power set.

Definition 1 (Test)
A test is a function $\varphi: \Omega \rightarrow \{0,1\}$, which associates a bit to each subset of Ω.
We concentrate on the following work:

Definition 2 (And-Tests)
An and-test $\varphi: \Omega \rightarrow \{0,1\}$ is a test which follows the following property:

$$\forall T \in \Omega \qquad \varphi(T) = \wedge\varphi(\{t\})$$

i.e. the result of an and-test on a given set is exactly equivalent to logical and of the test results on unit members of the given data set.

Remark 1. Note that 'or-tests' in which ∧ is replaced by ∨ in the definition, are exactly dual to our setting.
'xor-tests' are not investigated here. Although theoretically interesting by their right, we do not address the situation where both and-tests and or-tests are available, since we know of no concrete application where this is the case.
We can consider elements of Ω as n-bit strings, with meaning where the i-th bit indicates if i belongs to the subset. We say selection an element of Ω.

Definition 3 (Outcome).
The outcome $F\varphi(T)$ of a test φ on $T \in \Omega$ is the string of individual.
Test outcomes: $F\varphi(T) = \{\varphi(x), x \in T\} \in \{0,1\}^n$. When T = [n], Fφ will concisely represent Fφ ([n]).

Our purpose is to determine the outcome of a given test φ, by minimizing the expected number of queries to φ. Note that this minimized presupposition is trivially upper bounded by n.

Definition 4 (Splitting).

Let T ∈ Ω be a selection and φ be a test function. Let $\mathfrak{S}$ be a subset of Ω. The positive part of $\mathfrak{S}$,T, $\mathscr{S}_T^{\top}$, is defined as the set .

$$\mathscr{S}_T^{\top} = \{S | S \in \mathscr{S}, S \wedge T = T\}$$

where the operation ∧ $\mathscr{S}_T^{\perp} = \mathscr{S} - \mathscr{S}_T^{\top}$ is called the negative part of S with respect to T.

Definition5 (Testing procedure)

A testing procedure is a binary tree T which is labeled with nodes and leaves, such that:

1. The end nodes of T are in one-to-one correspondence with Ω in string representation; 2.
2. Each node of tree T has exactly two children, (S⊥, S>), labeled as (S, T) where S⊆Ω and T ∈Ω, such that

S⊥∩S> = 0
S⊥ tS > = S
S⊥ = S⊥ T and S> = S> T

Remark 2. It keeps to definition 5 that a testing procedure is always a finite binary tree, and that no useless φ are performed. It would give results as empty S for one of the children nodes. Furthermore, the root node has
S =Ω.

Pooling Procedure-

Consider a testing procedure T, outlined as on top. T describes the subsequent algorithm. On every node (S, T), perform the check φ on the selection T of samples. If φ (T) = 0, head to the left child; otherwise head to the right child. Note that on every node of a testing procedure, only 1 invocation of φ is performed.
The tree is finite and therefore this algorithm reaches a leaf S final during a finite variety of steps. By design,
S final = F φ.

Remark 3. Now, fix φ and assume it implicitly.

Probabilities on trees

To determine how efficient any given testing procedure is, we need to introduce a probability measure,
and a metric that counts how many calls to φ are performed.
We consider the discrete probability space (Ω, Pr). The expected output value of a random variable X is classically defined as:
E[X] = ∑ X(ω) Pr(ω)
Ω∈ Ω

Let T a testing procedure, and let S ∈ Ω be one of its leaves. The length ` T (S) of T over S is the distance on the tree from the root of T to the leaf S. This corresponds to the number of tests required to find S if S is the outcome of φ. The expected length of a testing procedure T is defined naturally as:
L T = E [` T] = ∑T (ω) Pr (ω) Ω ∈ Ω

It remains to specify the probabilities Pr(ω), i.e. for any given binary string ω, the probability that ω is
the outcome.

If the different tests are independent, we can answer this question directly with the following result:

Lemma 1. Assume that the events 'φ ({i}) = 1' and 'φ ({j}) = 1' are independent for i 6 = j. Then,
∀ω ∈ Ω, Pr(ω) can be written as a product of monomials of degree 1 in x 1. ... x n where x i = Pr(φ ({i}) = 1) = Pr (I th bit of ω = 1).
Thus, L T is a multivariate polynomial of degree n with integer coefficients.
In fact, or-tests provide inherently independent tests. Therefore, we will safely assume that the independence assumption holds.

Example 1. Let n = 5 and ω = 11101, then Pr (ω) = x 1 x 2 x 3 (1 − x 4) x 5

Remark 6. L T is uniquely determined as a polynomial by the integer vector of length 2 n defined by all its lengths: (T) = (T (0...0), ... T (1...1)).

Pool test optimization

As we introduced our approach. We try to optimize the testing procedures T (having smallest Lt)

Test Procedure:

We will generate sample cases for n>=1.
The objective is to implement a generation algorithm on the basis of testing procedure.

Algorithm 1: FindProcedure

Input: $S \in \Omega, C \in \Omega$.
Output: A binary tree.

1. if $|S| == 1$ then return S
2. $S'_{\perp} = S'_{\top} = C' = \emptyset$
3. for each $c \in C$
4. $S_{\perp} = S_c^{\perp}$
5. $S_{\top} = S_c^{\top}$
6. if $S_{\perp} \notin S'_{\perp}$ and $S_{\top} \notin S'_{\top}$
7. $S'_{\perp} = S'_{\perp} \cup \{S_{\perp}\}$
8. $S'_{\top} = S'_{\top} \cup \{S_{\top}\}$
9. $C' = C' \cup \{c\}$
10. for $i \in \{1, \ldots, |C'|\}$
11. $\overline{C} = C - C'[i]$
12. for each $\mathscr{T}_{\perp} \in$ FindProcedure$(S'_{\perp}[i], \overline{C})$
13. for each $\mathscr{T}_{\top} \in$ FindProcedure$(S'_{\top}[i], \overline{C})$
14. yield $(C'[i], \mathscr{T}_{\perp}, \mathscr{T}_{\top})$

IV.Conclusions

We have put forth the question of excellent pool testing with a priori probabilities, where one is given a set of samples and must regulate in the lowest average number of operations whose samples come out to be negative, and which are not. We formalized this problem and pointed out several interesting combinatorial and algebraic properties that speed up the computation of an optimal sequence of operations — which we call a meta procedure. We discovered the feasible solution for up to 4 samples. For larger values, our approach requires too much computation to be tractable, and thus an exact solution is out of reach; however, we gave several heuristic algorithms that scale well. We exhibit that these heuristics are sub-excellent in all cases, but they always do better than quality screening. The existence of a polynomial-time algorithm that finds optimal meta procedures for large values of n is an open question — although there is probably more hope in finding better heuristics. A substitute would be to change our generation algorithm to split branches when the outcome of expected lengths are all worse than any already-known procedure. Once the meta procedure for a given n is known, which only needs to be computed once, implementation is straightforward and only invokes a handful of (automatically generated) cases.

ACKNOWLEDGEMENT

This research was supported/partially supported by [King George Medical University, Lucknow]. We pay immense gratitude to every contributing body who helped with the valuable expertise that greatly impacted the quality, although they may not agree with all of the interpretations/conclusions of this paper.

We thank Dr. Indradeep Verma for assistance with guidance techniques, and resources from King George Medical University for availing the resources that greatly improved the manuscript.
We pay our deepest gratitude to Ankit Shukla for sharing their pearls of wisdom with us during the course of this research, and also, we are thankful to 3 “anonymous” reviewers for their so-called insights.

Our special thanks are extended to the staff of IIMT College of Engineering for their valuable support.

REFERENCES

1. Laboratory testing for coronavirus disease 2019 (COVID-19) for cases under suspicion. Interim Guidance WHO,02 March.
https://apps.who.int/iris/bitstream/handle/1

2. Global Surveillance for human infection with coronavirus disease (COVID-2019), Interim guidance, Geneva, World Health Organization,2020. .
https://www.who.int/publicationsdetail/global-surveillance-for-humaninfection-with-novel-coronavirus.

3. Managing epidemics, key facts about major deadly diseases. Geneva: World Health Organization,2018.
https://apps.who.int/iris/handle/10665/272442.

4. Laboratory biosafety guidance related to the novel coronavirus (2019-nCoV), World Health Organization, 2020.
https://www.who.int/docs/defaultsource/coronaviruse/laboratory-biosafetynovel-coronavirus-version-1-1.pdf?sfvrsn=912a9847_2